CS 134: Networks	Lecture 3, Monday, 1/30
Prof. Yaron Singer	Spring 2017

In the next two lectures, we will be interested in understanding how one can acquire information using a social network: given a piece of information held by an individual, how can we design algorithms that find this information efficiently?

Milgram's experiment. In 1967 while a graduate student at Harvard, Stanley Milgram performed an experiment in an attempt to verify the theory of a small world. The theory of a small world argues that every two people are separated by a short chain of acquaintances. To verify this theory, Milgram selected a set of people at random in the Midwest and asked them to forward an envelope to a specific person they did not know in Massachusetts. After some trial and error, Milgram launched an experiment where a good fraction of the envelopes reached their destination, in a mean number of steps of six, which coined the term six-degrees of separation.

What kind of graphs allow for these short paths to exist? What kind of graphs allow for people to be able to send information to strangers through acquaintances so efficiently?

1 Models of "Small World" Networks

Recall that a social network is modeled by a graph G where the set of nodes correspond to individuals and where edges represent how individuals are connected (e.g. friends). We are interested in graph models that are descriptive of social networks. In particular, we would like to understand the topological properties of graphs that explain phenomena like the one observed in Milgram's experiment.

Let's begin by surveying simple types of graphs as a first attempts to model social networks. These graphs give us some intuition of the different properties we should be looking for.

1. **Clique:** In a clique (see Figure 1), every node is connected to every other node in the graph. This model fails to capture our intuitive understanding of a social network, as the degree of every node is very large. We therefore wish to explore models in which every node has constant degree.

2. **Tree:** In a tree (see Figure 2), the graph has no cycles and every vertex is reachable from every other vertex. This model however fails to capture another intuitive feature of social networks and it is that some of our friends are likely to be friends themselves.

3. **Cycle:** In a cycle we have a path of n nodes in which the first node on the path is connected to the last node. The issue with this model is that it has high diameter (recall that the diameter of a graph is the largest distance between any two nodes in the network). As exemplified by Milgram's experiment, the distance between any two nodes of a social network should be low.

4. $G(n, p)$ **Random graph:** The $G(n, p)$ random graph is an appealing model. The problem however, is that in order for us to have small distances we need the average degree to be logarithmic. To see this, we can work out what the average degree of the random graph needs to be in order for us to have a connected graph with probability at least $1/n$.

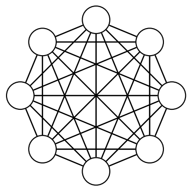

Figure 1: A clique.

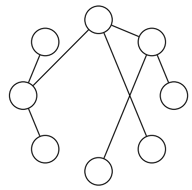

Figure 2: A tree.

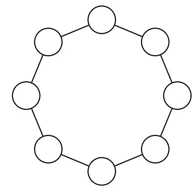

Figure 3: A cycle.

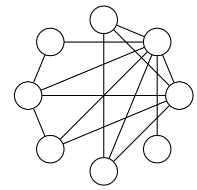

Figure 4: $G(8, 1/2)$.

We know how to measure the degree of nodes and the diameter of a graph, but how should we measure the notion that some of our friends are likely to be friends themselves? This notion is measured using the *clustering coefficient* from Assignment 2, which we define again as a reminder. Recall that $N(v)$ denotes the set of neighbors of v and $d(v)$ is the degree of node v.

Definition. *Given a graph $G = (V, E)$, the* ***clustering coefficient of a vertex*** *v, denoted $C(v)$, is the fraction, over all pairs of neighbors of v, of those pairs who are neighbors of each other. Formally,*

$$C(v) := \frac{|\{(u, w) \in E : u, w \in N(v)\}|}{\binom{d(v)}{2}}.$$

The ***clustering coefficient of a graph*** *is the average clustering coefficient of its nodes.*

Intuitively, if you were to pick two of your friends uniformly at random, then the probability that they are also friends is your clustering coefficient. For real-world social networks, we expect reasonably high clustering coefficients; ideally constant (that is, independent of the graph size). Observe that trees have a clustering coefficient of zero and that cliques have a clustering coefficient of one.

Properties of Social Networks:

- Small diameter,
- Large clustering coefficient, and
- Nodes should have a small degree.

When we examine the examples of graphs above, each graph has a property that defies our intuitive understanding of what social networks should look like. In the clique, every node has degree of $n - 1$. If we think of n being proportional to the size of the world population, having a model in which all nodes have such high degree seems unreasonable. In the tree example the clustering coefficient is 0; in your circle of friends, surely some of your friends are also friends amongst themselves, hence clustering coefficient of 0 again seems like an inappropriate model. Lastly, if we think about the ring, we can quickly see that it has diameter $n/2$, which is too large. In addition, it also has a clustering coefficient of 0.

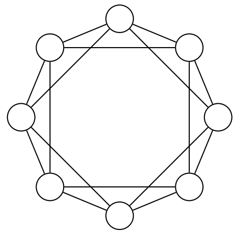

Figure 5: A ring lattice with $n = 8$ nodes and $k = 2$

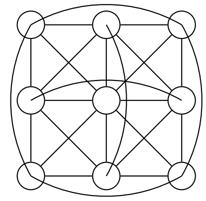

Figure 6: A square lattice with $r \times r = 3 \times 3 = 9$ nodes and $k = 2$.

2 Square Lattices, Expansion, and Regular Random Graphs

2.1 Square lattices have large clustering coefficient

Definition. *In a* ***square lattice*** *(see Figure 4) on $n = r^2$ nodes, the nodes are positioned in an $r \times r$ lattice and are connected to nodes at a lattice distance at most k away on the lattice for some given parameter $k \in \mathbb{N}$. The* ***lattice distance*** *between two nodes on a lattice is the minimum number of horizontal and vertical steps between the two nodes.*

A nice feature of square lattices is that they have high clustering coefficient; you will prove that on the next assignment for $k = 2$. Unfortunately lattices also have a large diameter that is linear in n. For social networks, the diameter of a graph should be at most logarithmic. In large social networks, say with one million or one billion nodes, $\log_2 n$ is roughly 10 and 20 respectively, which is reasonably small.

2.2 Regular random graphs have low diameter

Definition. *A d-regular random graph is a graph where all nodes have degree d and every two nodes have the same likelihood to be connected. That is, fix some number of vertices n, consider all graphs on n vertices that are d-regular, and pick one of these uniformly at random.*

For any $d \leq n - 1$, it is always possible to construct a d-regular random graph and we will talk more about that in a few lectures. We recall the *expansion* of a graph from assignment 2. [1]

The crucial property of d-regular is that they are likely to be good expanders. That is, a d-regular random graph is likely to have an expansion of α. A graph G is an α-expander if each vertex set $S \subset V$ with $|S| \leq n/2$ has at least $\alpha|S|$ edges leaving. Intuitively, this means that no two "large chunks" of the graph can be disconnected by removing few vertices or edges.

[1] More precisely, these are multi graphs, i.e. these are graphs that can also have self loops and multiple edges between the same pair of nodes; as we will see when we discuss these graphs, in these muligraphs, it is unlikely to have more than a negligible number of self loops and multiple edges.

Definition. *The* expansion *of a graph $G = (V, E)$ is the minimum, over all cuts we can make (dividing the graph in two pieces), of the number of edges crossing the cut divided by the number of vertices in the smaller half of the cut. Formally, it is*

$$\alpha = \min_{S \subseteq V, 1 \leq |S| \leq \frac{|V|}{2}} \frac{|e(S)|}{|S|}$$

where $e(S)$ is the number of edges leaving the set S.

So how does the expansion property help us? The following theorem shows that the expansion of a d-regular graph provides a guarantee on its diameter. Recall that a d-regular graph is a graph where all nodes have degree d.

Theorem 1. *Suppose a graph G is d-regular, for some constant $d \geq 3$, and has constant expansion α. Then the diameter of G is $O\left(\frac{d}{\alpha}\log n\right)$.*

Proof. We show that any two vertices s and t are a distance at most $O\left(\frac{d}{\alpha}\log n\right)$ apart. Let S_j be the set of vertices reachable from s in at most j steps (we can think of S_j as being formed by a breadth-first-search that starts at s).

Consider some j such that $|S_j| \leq n$. Because G has expansion α, there are at least $\alpha|S_j|$ edges leaving S_j. Since G is d-regular, there are at least $\alpha|S_j|/d$ vertices outside of S_j connected to S_j.

Therefore,

$$|S_{j+1}| \geq |S_j| + \frac{\alpha}{d}|S_j| = \left(1 + \frac{\alpha}{d}\right)|S_j|.$$

Because $S_0 = \{s\}$, we get

$$|S_j| \geq \left(1 + \frac{\alpha}{d}\right)^j.$$

Now pick $j = \frac{d}{\alpha}\log n$. Then we have

$$|S_{\frac{d}{\alpha}\log n}| \geq \left(1 + \frac{\alpha}{d}\right)^{\frac{d}{\alpha}\log n}.$$

We use the well-known fact that $\left(1 + \frac{1}{k}\right)^k \geq 2$ for $k \geq 1$ to get that $|S_{\frac{d}{\alpha}\log n}| \geq 2^{\log n} = n$. [2]

Therefore, the size of S_j reaches at least $\frac{n}{2}$ before this point, *i.e.* before $j = \frac{d}{\alpha}\log n$. Now, by the exact same reasoning, if we start at t and consider the sets T_j, we find that the size of T_j reaches at least $\frac{n}{2}$ before $j = \frac{d}{\alpha}\log n$. But then $T_{\frac{d}{\alpha}\log n}$ and $S_{\frac{d}{\alpha}\log n}$ must have some vertex in common (since both are larger than $\frac{n}{2}$). That means there is a path from s to t of length at most $2\frac{d}{\alpha}\log n$, because we can go from s to this common vertex in at most $\frac{d}{\alpha}\log n$ steps, and then similarly to get from this vertex to t. □

So the above theorem implies that for $d \geq 3$, a d-regular graph with constant expansion α gives us the small world property that we want: it has small (logarithmic) diameter. Using randomness, we can construct d-regular graphs with constant expansion.

For $d \geq 3$, it turns out that d-regular random graphs have expansion that is a function of d only, not of the number of vertices (we skip the proof). While d-regular random graphs have small diameter, they unfortunately have low clustering coefficients (about d/n and we skip the proof as well).

[2]Note: This assumes that $k = \frac{d}{\alpha}$ is bigger than 1. But if it is less, then we can fix the argument by taking $j = c\frac{d}{\alpha}\log n$ with $c = \alpha/d$. Since α and d are constant, c is constant and the asymptotic result remains the same.

3 Constructing a Small World Model

Lattices have a large clustering coefficient but their diameters are too large. On the other hand, d-regular random graphs have small diameters but their clustering coefficients are too small. To get the best of both worlds we can simply combine the lattice with a d-regular random graph, for some constant $d \geq 3$. By that we mean that we will take a d-regular random graph and add the edges of a lattice to the nodes of the random graph. The result will be a graph with constant degree that has both small diameter and large clustering coefficient. This model is inspired from the Watts-Strogatz model [1], which is also composed of a graph with large clustering coefficient and of random edges.

The Watts-Strogatz model:[3]

- Nodes are positioned on a square lattice with n nodes (for simplicity, assume $\sqrt{n}$ is an integer).
- Each node is connected to the nodes that are most a distance k away on the lattice, for some given parameter $k \in \mathbb{N}$.
- For some given $\ell \in \mathbb{N}$ each node has ℓ *long range* edges which are connected uniformly at random to other nodes in the network.

Summing up, we obtain the following table.

	diameter	clustering coefficient	maximum degree
Lattice	bad (too large)	good (constant)	good (constant)
d-regular random	good (logarithmic)	bad (too small)	good (constant)
Watts-Strogatz	good	good	good

[3] In the Watts-Strogatz model, the square lattice is replaced by a ring lattice. However, the intuition remains the same as a ring lattice also has large diameter and constant clustering coefficient.

References

[1] Duncan J. Watts and Steven H. Strogatz. *Collective dynamics of small-world networks.* Nature, 393:440442, 1998.